White Paper on VRS

Contact: Ed Bosson | VP of Regulatory Affairs | ed@convorelay.com

The Theater of Video Relay Services

Video calling has brought new life to Telecommunications Relay Services (TRS), which was established by Congress under Title IV of the Americans with Disabilities Act.

Video Relay Services (VRS) allows deaf, hard of hearing, and speech impaired people to videoconference with sign language interpreters specially trained to relay phone conversations.

VRS users depend on the service for everyday communication; however, they face technology issues. This is because starting and establishing a video call has many challenges:

“What ports can we use?”
“Which video codec should we use?”
“Should we use TCP or UDP?”

Understanding the basic technologies involved in each video call helps us understand what VRS consumers experience and can lead to well-defined, well-written regulations that benefit VRS consumers.

The Actors in a Video Connection

You make a VRS call using a **DEVICE WITH A WEBCAM**. This might be a videophone, computer or a smartphone. *Imagine you are driving a vehicle – it could be a car, a truck, or a motorcycle.*

Your device uses a specific **PROTOCOL** to send and receive video. *Your vehicle uses either regular gas or diesel.*

The video is compressed for faster transmission; the technology for this is a **CODEC**. *You take one of several highways, all of which are paved differently.*

The video call goes to an **IP ADDRESS**. *This is the building address of where you are visiting.*

Each IP address has many **PORTS**. *Imagine the building has several doors.*

The Supporting Cast of a Video Connection

Protocols

In the early days of videoconferencing developers had their own method of sending and receiving video, also known as a protocol. So, devices that used different protocols would not connect with one other.

Today there are two major standards – H.323 and SIP:

- The International Telecommunications Union (ITU), a United Nations agency that regulates communication technology issues, endorsed H.323 - which is an umbrella for many sub-specifications. H.323, originally designed for Local Area Networks ("LANs") is comprehensive and inflexible.

- The Internet Engineering Task Force (IETF), a worldwide Internet standards organization, defined a new protocol called Session Initiation Protocol (SIP). SIP is not an umbrella specification but a complete specification designed and optimized to handle real-time multimedia content, such as voice and video calls.

H.323 and SIP accomplish the same goals in different ways; however, proponents of SIP have their roots in the Internet community rather than the international community.

H.323	H.323 listens on TCP port 1720 to initiate a video call. If blocked, a call cannot be established. When a call is detected through port 1720, the endpoints negotiate which other port(s) to use to communicate, typically between 1024 and 4999. If these ports are blocked the calling terminal may indicate that the other endpoint is ringing, but no connection will take place. Finally, ports in the range of 10,000 to 20,000 are used for video data. If these ports are blocked, the displays may show no video on one or both endpoints.
SIP	SIP listens on port 5060 and port 5061 if a private (encrypted) call is requested. This port is required to signal an invite and initiate a call. SIP partners with Session Description Protocol ("SDP") which is used to figure out how video will be transmitted between the two endpoints - which transport to use (TCP or UDP), which video codecs to use, and which ports to use. If there are no available ports to use, then video can't be transmitted, and the call will disconnect. In other words, SIP by itself gets things going; SDP does the actual exchange of information necessary to establish a call. Because of this, it is much more flexible.

H.323

SIP

Codecs

The word “codec” combines parts of the words ‘CO’mpression and ‘DEC’ompression and is, simply, a method of video compression. Compressed video requires less bandwidth to transmit.

H.261	The first video codec, became an ITU standard in 1990. Dial-up Internet connections were used to transmit H.261 video. H.261 has a very low frame rate and offers limited video resolutions. It is still offered as a standard in Sorenson videophones.
H.263	The second-generation codec, became an ITU standard in 1995 and is capable of much higher data rates – up to 30 frames per second. Its quality is superior to all prior standards.
H.264	Became an ITU standard in 2003, is the third generation and compresses video at twice the speed of H.263 while maintaining the same picture quality.
H.265	Currently under development, is expected to reduce the bit-rate required for video transmission by half. If achieved, this will increase usability of signing in low light situations, or high contrast situations where the hands are not distinguishable.

IP addresses

An Internet Protocol (IP) address functions as a unique destination, like a home address so the Post Office knows exactly where to send your mail. However, IP addresses are hard to memorize and can change between calls.

The solution is to register the videophone itself with a central server, like the iTRS database. This allows each VRS callers to give out a friendly 10-digit phone number instead of an IP address. When someone calls your phone number, the iTRS database automatically retrieves the information registered with your phone number then redirects the call to your videophone.

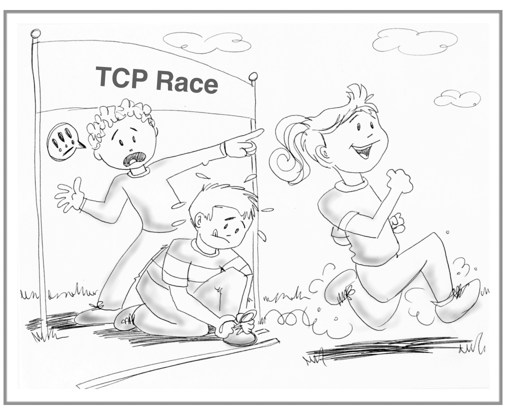
Ports

Your IP address has a range of ports (1 to 65,535) – in other words, “doors” of your home. Ports make video calls possible; an Internet without ports is like a worldwide network of houses with no doors, presenting no way to establish a video call.

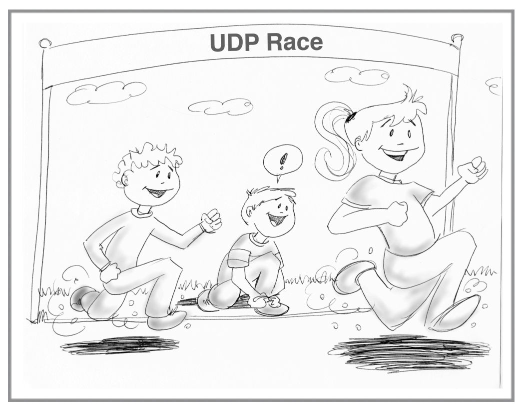
Ports are divided into, and use, two specific frame protocols: TCP and UDP.

- *Transmission Control Protocol (TCP): Designed and optimized for accuracy and has built-in error correction. TCP's tradeoff is that it creates additional network demands and slows down transmission. This makes TCP less than ideal for transmitting real-time video.*

- *User Datagram Protocol (UDP): Designed for uses where accuracy of transmitted data defers to speed of transmission. UDP's tradeoff is the loss of some data so the bulk of the transmitted data arrive at its destination as close to real time as possible – an important consideration for streaming video.*



TCP



UDP

Both frame protocols enable video connection between equipment and software developed by different companies. However, under circumstances of small network errors, TCP would completely halt the transmission of the video stream until the errors are compensated then transmit the backed-up data all at once to its destination. This results in – in visual terms – the video pausing then playing back at extremely high (and unintelligible) speeds until it returns to real-time.

UDP is the ideal choice because it is more flexible and prioritizes real-time transmission of the video stream. UDP, under normal circumstances, compensates for small network errors by ignoring them in favor of sending as much of the intact video stream to the destination as possible. While this sometimes results in pixilation and brief pauses/blackouts, in visual terms the video plays back mostly uninterrupted.

Behind the Curtains of a Video Connection

Connection, Negotiation and Streaming

When you call another person, what typically happens is that you use a device and dial a number, and the other person answers your video call on his/her device. You both can see each other, and chat throughout the phone call. A breakdown is as follows:

1. *Call connection: Your device connects to the other person's device. The phone is picked up!*

2. *Negotiation: The devices decide how video will transmit (also known as "signaling" or a "handshake protocol")…*

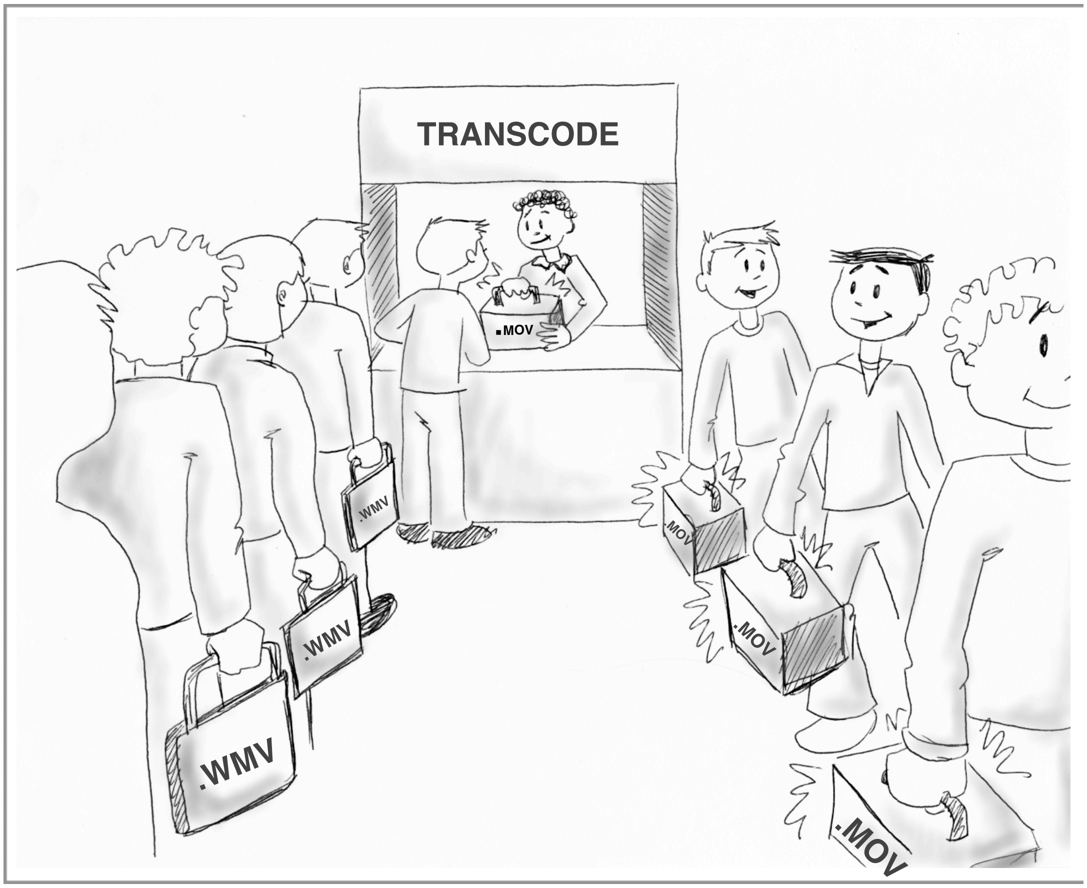
   "What ports can we use?"
   "Which video codec should we use?"
   "Should we use TCP or UDP?"

3. *Video connection: The video streaming begins. You see the other person and the other person sees you!*

When you call someone without an interoperable device, you will either not connect at all or you will connect but get a black screen, meaning there is no video connection.

Transcoding

For video to be clear enough for ASL conversations, the frame rate must stay above 15 frames per second, otherwise your finger spelling may not be legible; approximately 20 frames per second is the preferable minimum.



All this depends on the video codecs used.

Media gateways are servers that act as translators between differing codecs. Similar to VRS communication assistants, who take information communicated in sign language and interpret it into English and vice versa, media gateways are what makes video connection between equipment using different codecs possible.

In other words, media gateways transcode video data. For example, a media gateway can successfully translate between a H.323 endpoint that compresses its video using H.263, and a SIP endpoint that uses H.264.

Because of this capability, media gateways can bridge devices from different manufacturers and facilitate interoperability.

Your Mock Video Call, in Detail

You and your friend Pat want to videochat. You dial Pat's phone number: 800-555-1234. (Remember, the phone number is simply a friendly mask for an IP address.)

Imagine you are a packet, and you are carrying a briefcase that contains video data. You need to deliver it to "the ABC Room" (the friendly name) at the hotel. You arrive and the receptionist (the iTRS database) tells you that the ABC Room today is at room B-21315 (the IP address).

Your call to 800-555-1234 goes to the iTRS database, which automatically sends your call to Pat's device. Pat is excited and clicks on the "Accept Call" button.

Your and Pat's devices connect. Your videophone uses SIP, handshakes with Pat's videophone, and uses SDP to ask:

> *Hey buddy! 1) Which video codec are you using? 2) Which ports do you have open, and are they TCP or UDP? 3) How much bandwidth is available?*

Pat's videophone responds:

> *Hello! The same old same… H.264, 'baseline profile,' ports 80, TCP, and 5,000-9,999, both TCP and UDP, and 512 kilobytes per second.*

Your videophone now knows what to do so it can stream video, and starts by sending to port 5,000 – but only a limited number: 512 kilobytes per second. If there's a problem, port 5,001 is next… and so on, until there's an open port.

You arrive at the room but there are 5,000 doors – numbered 5,000 through 9,999. The room is a representation of Pat's IP address; the doors are representations of the ports. At first, you will bang on the nearest door – 5,000 – and if nobody opens it, you'll try the next door, and so on. The more doors you bang on, the more energy you use and the longer you take before someone finally lets you in.

Your videophone finds an open port, and Pat's videophone now has begun to receive video. However, Pat's videophone still needs to decompress the video stream.

Imagine you bang on a door and someone finally opens it. You enter – the open port – and the person who let you takes your briefcase then uses a key to open your briefcase.

Success! Pat can see the video you are sending! You two have a video connection!

But… suppose that you are using Flash while Pat is using H.264? The solution would be to use a media gateway. As above, your SIP videophone would use SDP to ask Pat's videophone the following questions:

> *Hey buddy! 1) Which video codec are you using? 2) Which ports do you have open, and are they TCP or UDP? 3) How much bandwidth is available?*

Pat's videophone would again respond:

> *Hello! The same old same… H.264, 'baseline profile,' ports 80, TCP, and 5,000-9,999, both TCP and UDP, and 512 kilobytes per second.*

This negotiation process – called the “signaling” or “handshake” protocol – gathers information from both endpoints and then uses that information to figure out what to do. If the codecs used at either end do not match, the negotiation process will communicate with the media gateway to step in, and both endpoints will stream their data through the media server. The end result is that the media gateway ensures that both endpoints receive video using the correct codec.

You leave the room. The person holding your briefcase looks inside and says,

> ***This briefcase is Flashy! I cannot deliver it like this!***

As a media gateway would, this person takes out the video data, stuffs it in a purse labeled H.264, and locks it with one of many, many keys hanging on his keychain. This person brings the purse to another person, who has a copy of the key that can open the purse.

This entire process is hidden from you. The transcoding of your Flash video into the H.264 format so that Pat can see you is completely automatic and you have no control over it, nor are you aware that it is happening.

Next Act

SIP and H.323 both make video calls possible, but SIP is superior

SIP has become the recognized standard for real-time audio and video. All videophones released in recent years are SIP-capable, and the new generation of video-capable mobile endpoints (iPhone4, HTC Evo 4G, etc) utilize SIP as well.

However, the use of the iTRS database needs to be expanded to accommodate videophone addresses beyond H.323. To date, SIP has received widespread support from the VRS industry, and should be adopted due to its superior flexibility.

Which video codec should VRS use?

Media gateways, properly used and configured, allow devices that use different video codecs to interoperate. With this uniformity in handshake protocols, the remaining major barrier to more universal interoperability between endpoints is video codecs.

You do not have to worry about "matching" your videophone to the equipment used by your friends and family. In addition, advanced products such as Web-based Flash video, which typically are not friendly to older video equipment, can be made interoperable while still bringing benefits to you.

However, for ideal video quality, specifically in the VRS context, a standard video codec should be chosen to minimize the number of video streams that are forced to go through a media gateway for compatibility. Using a standard video codec would also lower costs and bandwidth while the quality of video streams would remain constant across different devices and platforms. Media gateways, in turn, can still be used to enable newer and more advanced products to be backwards-compatible if a more advanced codec is chosen as a new industry standard at some point in the future.

There is precedent in establishing technological standards. For instance, the IEEE, a standards-setting body, provides a forum by which companies can collaborate to create industry standards. Wi-Fi is one such standard that came about as a result of this collaboration. In the VRS industry, similar collaboration can easily produce a VRS standard video codec.

About Convo

Convo Communications is deaf-owned and provides 24/7 video relay services, which allow phone conversations to and from sign language users living in a global economy. www.convorelay.com

Organic, natural VRS

Convo strives to provide a "natural VRS" experience, with the conversation flowing as if you were having a direct conversation with the person you are calling, completely free of technology and interpreter snags.

Convo pledges to run its business ethically and to not pump its bottom line with calls that are intended only to create revenue. To this end, Convo handles only organic calls -- that is, VRS calls that are legitimate.

Some Facts About Us

Headquarters	San Ramon, CA
Locations	Bay Area, CA; Sacramento, CA; Austin, TX; New York, NY; Seattle, WA; Mobile, AL
Partners	SignOn Sign Language Services
Products	Convo Anywhere, Green Book, ConvoIM
Incorporation	Limited Liability Company
Convo is	• 100% deaf-owned • Open 24/7 year-round • Privately owned and without funding from equity groups • With employees, for whom 95% actively use sign language • Among the top five largest VRS providers

March 2009, Convo is born

Mar 2009	Convo founded
May 2009	Launches relay service
Aug 2009	Launches ConvoIM
Jan 2010	Opens Seattle call center in partnership with Sign On Language Services
Feb 2010	Moves HQ to San Ramon, CA Launches Convo Green Book Opens San Ramon call center
Apr 2010	Opens Mobile, AL call center
Jul 2010	Opens Roseville call center
Oct 2010	Launches Convo Anywhere
Oct 2010	Revamps website